

AN RDF REPRESENTATION FOR SEMANTIC MUSIC PRODUCTION DATA

Sean Enderby and Ryan Stables

Digital Media Technology Lab
Birmingham City University
sean.enderby@mail.bcu.ac.uk
ryan.stables@bcu.ac.uk

Thomas Wilmering and György Fazekas

Centre for Digital Music
Queen Mary University of London
{t.wilmering, g.fazekas}@qmul.ac.uk

ABSTRACT

Music production involves a wide range of techniques that can be described by natural language. Capturing these descriptions at the source allows us to understand the intentions of an engineer and consequently develop intelligent tools and interfaces by computationally modelling them. In this paper we present a database architecture for capturing these attributes in a digital audio workstation, in which we retain audio features, audio-effect parameters, user informatics and semantic descriptions of the audio transformations. This allows us to build a comprehensive map of the audio engineering workflow using linked-data, which can be utilised on the semantic web. We show that attributes such as provenance, which is omitted from relational database models can be a useful indicator of data validity.

1. INTRODUCTION

In music production, natural language is often used to describe timbral transformations. Recently, these descriptions have been the focus of intelligent music production research [1], as they allow for the development of systems that provide intuitive control of technical processes. To facilitate this, we present a model for the representation of semantically annotated music production data.

2. THE SAFE ONTOLOGY

The SAFE Ontology¹ is an extension of the Studio [2] and Audio Effects [3] Ontologies, designed to represent the application of audio effects in music production and the semantic descriptions thereof. The data is gathered using the SAFE audio plug-ins [4], and comprises:

- Details of the processing applied to a signal (which audio effect used and its parameter settings).
- A semantic description of the timbral effect of the processing.
- Audio features of the signal before and after processing.
- Metadata about the signal (instrument, genre).
- Metadata about the processing (location).

¹Available at <http://www.semanticaudio.co.uk/datasets/safe-rdf>

- Metadata about the user (age, primary language, production experience).
- Provenance of the above data (how the data was produced e.g. human input or computer analysis).

2.1. Transform Data

Each entry in the SAFE dataset is described using the *studio:Transform* concept. They all apply some transform to a set of input signals to produce a set of output signals. Each is given a semantic label, its *safe:DescriptorItem*, describing the timbral effect it had on these signals. In the SAFE Ontology the transforms are described using the set of RDF triples shown in Figure 1.

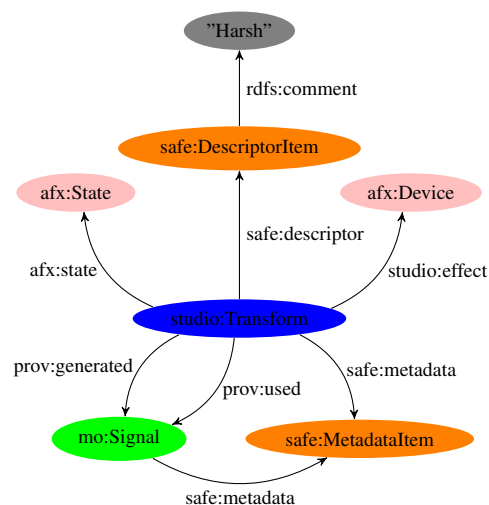


Figure 1: The structure used to describe the application of an audio effect.

Metadata items are used to provide details about the application domain of the effect. Each *safe:MetadataItem* describes one property of an object, the property being identified using an *rdfs:label* and the description using an *rdfs:comment*. Each object described by metadata has its own set of properties, “genre” and “instrument” metadata tags, for example, describe an audio signal (*mo:Signal*), while “location” tags describe a transform.

2.2. Audio Feature Data

The analysis of an audio signal is described using the *safe:FeatureExtractionTransform* concept. This is similar

to the *studio:Transform* concept but uses an audio signal to generate a time series of feature values. Every signal used by a transform has its own set of feature extraction transforms which describe it. The temporal locations of each feature value within a signal are described through use of the Timeline Ontology. Audio features are taken from both the input and output signals to aid in semantic analysis. Patterns found in the audio features suggest that a term describes the output signal of a transform, whereas patterns found in the change in audio features between input and output signals suggest a term describes the effects of the transform itself. This is highlighted in Figure 2

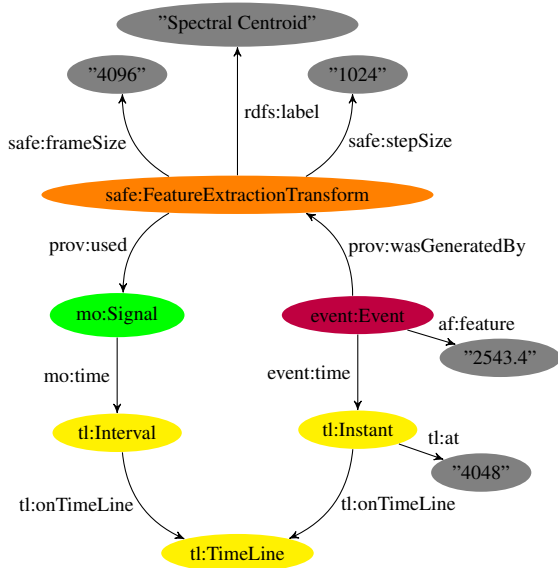


Figure 2: The structure used to describe the features of an audio signal.

3. PROVENANCE DATA

The SAFE Ontology makes extensive use of the Provenance Ontology to record the origins of the various data. The interface of the SAFE plug-ins requires that the user provide a semantic description of their use of the plug-in. Therefore the provenance of every *safe:DescriptorItem* is attributed to the user who saved it. It is, however, not mandatory for users to fill in the metadata fields. These can be later populated through analysis of the audio features of the signal. The provenance of the metadata items provides a method to distinguish between the more reliable user submitted metadata and the less reliable computer generated metadata.

Missing metadata is estimated using a collaborative filtering technique commonly associated with recommender systems [5]. The reliability of this computer generated metadata is then compared against that provided by users. Firstly, redundant data is removed by applying principal component analysis to the audio feature data associated with each signal. The first 10 components are retained, describing over 97% of the total variance. The reliability of the metadata is measured as the mean within-class variance across the first 10 principal components for each metadata tag. The results of this are shown in Table 1.

| Tag | Genre | | Instrument | | |
|---------------|---------|---------|------------|---------|---------|
| | var_h | var_m | Tag | var_h | var_m |
| blues | 0.031 | 0.035 | bass* | 0.024 | 0.032 |
| classical* | 0.027 | 0.031 | drums* | 0.008 | 0.033 |
| electronica | 0.002 | 0.004 | guitar | 0.026 | 0.017 |
| experimental* | 0.011 | 0.188 | hi-hat* | 0.000 | 0.014 |
| funk | 0.026 | 0.021 | kick* | 0.015 | 0.023 |
| jazz* | 0.014 | 0.020 | organ | 0.025 | 0.026 |
| metal | 0.027 | 0.014 | piano* | 0.031 | 0.095 |
| pop | 0.015 | 0.019 | snare | 0.028 | 0.022 |
| reggae | 0.050 | 0.040 | trumpet* | 0.045 | 0.072 |
| rock | 0.145 | 0.009 | vocals | 0.045 | 0.011 |
| mean | 0.035 | 0.038 | mean | 0.025 | 0.035 |

Table 1: The variance across the first 10 Principal Components for each of the tags in the genre and instrument categories using human labelled (var_h) and machine labelled (var_m) metadata entries, * represents a significant increase in variance ($p < .005$).

The results show that the machine labelled entries exhibit .001 higher variance than those labelled by humans for genre and .01 higher variance for instrument classes. Whilst only a subset of the tags exhibit a statistically significant increase in variance, this suggests the provenance ontology plays an important role in describing the reliability of both instrument and genre tags.

4. REFERENCES

- [1] R. Stables, S. Enderby, B. De Man, T. Willmering, G. Fazekas, and J. Reiss, “Semantic description of timbral transformations in music production,” in *ACM conference on Multimedia*, 2016.
- [2] G. Fazekas and M. B. Sandler, “The studio ontology framework,” in *12th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 471–476, 2011.
- [3] T. Wilmering, G. Fazekas, and M. B. Sandler, “The audio effects ontology,” in *14th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 215–220, 2013.
- [4] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. Reiss, “Safe: A system for the extraction and retrieval of semantic audio descriptors,” in *15th International Society for Music Information Retrieval Conference (ISMIR)*, 2014.
- [5] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 8, pp. 30–37, 2009.